Anal. Chem. **2003**, *75*, 3817−3822

Microorganism Identification by Matrix-Assisted Laser/Desorption Ionization Mass Spectrometry and Model-Derived Ribosomal Protein Biomarkers

Fernando J. Pineda,*,† Miquel D. Antoine,‡ Plamen A. Demirev,‡ Andrew B. Feldman,‡ Joany Jackman,‡ Melissa Longenecker,‡ and Jeffrey S. Lin‡

Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205, and Research and Technology Development Center, Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723-6099

An improved data analysis method is described for rapid identification of intact microorganisms from MALDI-TOF-MS data. The method makes no use of mass spectral fingerprints. Instead, a microorganism database is automatically generated that contains biomarker masses derived from ribosomal protein sequences and a model of N-terminal Met loss. We quantitatively validate the method via a blind study that seeks to identify microorganisms with known ribosomal protein sequences. We also include in the database microorganisms with incompletely known sets of ribosomal proteins to test the specificity of the method. With an optimal MALDI protocol, and at the 95% confidence level, microorganisms represented in the database with 20 or more biomarkers (i.e., those with complete or nearly completely sequenced genomes) are correctly identified from their spectra 100% of the time, with no incorrect identifications. Microorganisms with seven or less biomarkers (i.e., incompletely sequenced genomes) are either not identified or misidentified. Robustness with respect to variations in sample preparation protocol and mass analysis protocol is demonstrated by collecting data with two different matrixes and under two different ion-mode configurations. Statistical analysis suggests that, even without further improvement, the method described here would successfully scale up to microorganism databases with roughly 1000 microorganisms. The results demonstrate that microorganism identification based on proteome data and modeling can perform as well as methods based on mass spectral fingerprinting.

Rapid and reliable identification of microorganisms is of paramount importance for advancing homeland security.[1] Matrix-assisted laser desorption/ionization − time-of-flight (MALDI-TOF) mass spectrometry (MS) is emerging as a technology capable of fulfilling this task. This technology generates mass spectra with unique biomarker profiles on a time scale of minutes from intact microorganisms, with very minimal sample preparation.[2,3]

The prototypical approach with this technology is to identify a microorganism from its experimental mass spectrum by measuring the similarity between its spectrum and the mass spectra in a reference ("fingerprint") library. Correlation coefficients,[4] root-mean-square differences,[5] and statistical *p*-values[6] are commonly used to quantify spectral similarity. A high degree of reproducibility is required for such fingerprint approaches to be effective. But mass spectral reproducibility is sensitive to variations in interlaboratory protocols and mass spectrometer settings.[2,7] Consequently, fingerprint approaches are constrained to use identical sample preparation protocols, instruments, and settings. In addition to variability due to sample preparation, biochemical processes in microorganisms contribute to the variability of MALDI mass spectra.[2,8,9]

To perform rapid and robust identification in the face of mass spectral variability, an approach for microorganism identification based on proteome database queries was recently proposed.[9] A hypothesis test was subsequently introduced to quantify the significance of these identifications.[10] The test statistic of this hypothesis test is a *p*-value that estimates the probability of misidentification due to accidental matches between experimental peaks and database proteins of unrelated microorganisms. The *p*-value reflects the probability of obtaining the observed number of matches by chance alone. Thus, the lower the *p*-value, the less likely it is that the matches occurred by chance. Accordingly, lower *p*-values correspond to more significant identifications. The *p*-value accounts for the mass accuracy, the biomarker density

* To whom correspondence should be addressed. E-mail: fernando.pineda@jhu.edu.

† Department of Molecular Microbiology and Immunology.

‡ Research and Technology Development Center.

(1) Committee on Science and Technology for Countering Terrorism. NRC. *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism*; National Academy Press: Washington, DC, 2002.

(2) Fenselau, C.; Demirev, P. *Mass Spectrom. Rev.* **2001**, *20*, 157−171.

(3) Lay, J. O., Jr. *Mass Spectrom. Rev.* **2001**, *20*, 172−194.

(4) Arnold, R. J.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 630−636.

(5) Claydon, M. A.; Davey, S. N.; Edwards-Jones, V.; Gordon, D. *Nat. Biotechnol.* **1996**, *14*, 1584−1586.

(6) Jarman, K. H.; Cebula, S. T.; Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Kingsley, M. T.; Wahl, K. L. *Anal. Chem.* **2000**, *72*, 1217−1223.

(7) Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 456−464.

(8) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105−112.

(9) Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732−2738.

(10) Pineda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000**, *72*, 3739−3744.

10.1021/ac034069b CCC: $25.00
Published on Web 06/28/2003

(the number of database proteins per unit mass interval for a given microorganism), the number of experimental mass spectral peaks submitted to the search, and the number of these peaks that match a microorganism's biomarkers.[10]

Recent analysis of mass spectra from intact *Helicobacter pylori* demonstrated that the identification significance (as measured by *p*-values) could be improved by 1 order of magnitude, if the search were modified to account for N-terminal Met loss—a posttranslational modification (PTM) that commonly occurs in prokaryotes.[11] This PTM reduces the molecular mass of a modified protein by 131 Da.

While an important step forward, the proteomics-based approach described above does not yield a practical assay for microorganism identification because the *p*-values obtained in these studies were not sufficiently low to be considered statistically significant identifications. Essentially, this is due to the fact that naïve proteome database search treats all the proteins in the proteome of a microorganism as biomarkers. Most of these proteins have a low a priori probability of being observed. This is reflected by the fact that the number of peaks (10−30) in a typical MALDI mass spectrum of an intact cell is much less than the typical number of proteins (500−4000) found in a microorganism's proteome. Such discrepancy implies that directly comparing mass spectral peaks to the entire proteome of a microorganism is likely to result in a large number of false matches. This can only reduce the significance of identifications. Here we demonstrate that by accounting only for the most abundantly expressed proteins in vegetative cells (e.g., ribosomal proteins), we can reduce the number of false matches and thereby improve the significance of identifications by several orders of magnitude. This yields, for the first time, a practical automated proteomics-based MALDI-TOF assay for rapid microorganism identification.

EXPERIMENTAL PROTOCOL

Test Organism Cultures and Sample Preparation. Organisms were isolated on tryptic soy agar and incubated overnight at 37 °C (or 55 °C, depending on organism growth conditions). One colony from each plate was inoculated into 5 mL of tryptic soy broth (TSB) and incubated overnight at 37 °C (or 55 °C), 100 rpm in a shaking incubator. One milliliter from the culture was further inoculated into 4 mL of TSB and incubated 6 h at 37 °C (or 55 °C), 100 rpm in a shaking incubator until log-phase growth for optimal ribosomal protein expression was obtained (monitored at 600 nm). One milliliter (10^8 cfu/mL) of each organism was then frozen at −80 °C. *Haemophilus influenzae* was grown on chocolate agar and incubated overnight at 37 °C, 5% CO_2. One colony was removed from the plate and inoculated into TSB for a 10^8 cfu/mL concentration of the organism. The organism was then frozen at −80 °C. After harvesting, each culture was assigned a coded label. Samples were prepared for MALDI analysis by thawing, and the culture medium was washed with either water or 2% ammonium chloride. Pelleted bacteria were resuspended in water, and 0.5 μL was deposited onto a stainless steel slide, followed by addition of 0.5 μL of matrix solution. Either 150 mM α-cyano-4-hydroxycinnamic acid (CHCA) saturated sinapinic acid (SA) in acetonitrile/water (5% trifluoroacetic acid), 70:30 (v/v) was used as matrix. Equine cytochrome *c* was also added to provide an internal standard for mass calibration.

Mass Spectrometry. Positive or negative ion spectra were acquired in linear mode on a Kompact MALDI Discovery (Kratos Analytical Instruments, Chestnut Ridge, NY) TOF instrument. The nominal accelerating voltage was ±20 kV. The N_2 laser (337 nm) had an estimated fluence of 10 mJ/cm² before attenuation. Pulsed ion (delayed) extraction was optimized for ion focusing and transmission at m/z 10^4. The estimated mass accuracy was ±5 Da (and used in the identification algorithm as well). Each spectrum was the average of 50 consecutive laser shot traces, with the beam rastered linearly across the entire sample well. Several replicate spectra were taken for each sample under the same experimental conditions. The five replicate spectra that had highest intensity signals were selected to represent each sample. Peak lists were extracted with the software provided with the instrument, including only peaks with amplitude above 2 mV. The peak lists were compiled by assuming that only singly charged (protonated or deprotonated) molecular ions were detected (not always valid, vide infra). The cytochrome *c* calibrant ions were removed from the peak lists.

RESULTS

Microorganism Biomarker Models. We refer to a ribosomal protein as a *ribosomal protein biomarker* or simply as a *biomarker*. We refer to a microorganism's set of automatically generated ribosomal protein biomarkers as a *microorganism model*. Modeling starts with protein sequence data from the SWISSPROT (Rel. 39.7) and TrEMBL (Rel. 14.17) databases.[12] In this study, we restrict ourselves to ribosomal proteins. Ribosomal proteins are highly abundant in vegetative cells—up to 20 wt %, relative to other cytosolic proteins[8,13]—and are easily selected from SWISSPROT/TrEMBL by a query for the term "ribosomal" in the "DE" field of a SWISSPROT record. Our microorganism models do not include ribosomal protein fragments, proteins that contain ambiguous residues (i.e., "amino acids" B, X, or Z), or proteins outside the 4−13-kDa mass range. The latter reflects the observation that most protein biomarkers in MALDI mass spectra from intact microorganisms are within that range.[2,3] Only 18 microorganisms in the database have 20 or more ribosomal protein biomarkers. *Escherichia coli* has the most—31 biomarkers. Finally, microorganisms represented by less than three biomarkers are excluded from the database. Although there are hundreds of microorganisms represented in the original SWISSPROT/TrEMBL database, the modeling process described above significantly reduces the database size to 38 microorganism models. Table 1 lists the microorganisms in the database along with the number of biomarkers used to represent each microorganism. The broad range in the number of biomarkers in the models reflects the completeness of their respective genome-sequencing projects and their state of annotation, rather than the actual number of ribosomal proteins in the microorganisms.

To automatically account for N-terminally Met-cleaved sequences, we apply a deterministic rule for N-terminal Met loss to all biomarker sequences. According to this rule, the N-terminal Met residue is automatically cleaved, or remains intact, depending

(11) Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566−4573.

(12) O'Donovan, C.; Martin, M. J.; Gattiker, A.; Gasteiger, E.; Bairoth, A.; Apweiler, R. *Brief Bioinform.* **2002**, *3*, 275−284.

(13) Ryzhov, V.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 746−750.

Table 1. List of the Microorganisms Included in the Database and the Number of Ribosomal Proteins in the 4–13-kDa Mass Range[a]

n	cultured	species
31	√	*Bacillus subtilis*
30	√	*Escherichia coli*
26	√	*Pseudomonas aeruginosa*
25	√	*Haemophilus influenzae*
24		*Borrelia burgdorferi*
24		*Deinococcus radiodurans*
23		*Mycoplasma genitalium*
23		*Mycoplasma pneumoniae*
22		*Chlamydia pneumoniae*
22		*Chlamydia trachomatis*
22		*Helicobacter pylori*
22		*Helicobacter pylori* J99
22		*Mycobacterium tuberculosis*
22		*Rickettsia prowazekii*
21		*Treponema pallidum*
20	√	*Bacillus stearothermophilus*
20		*Synechocystis* sp
20		*Thermotoga maritima*
18		*Aquifex aeolicus*
16		*Streptomyces coelicolor*
15		*Thermus aquaticus*
14		*Mycobacterium leprae*
10		*Bacillus halodurans*
9		*Leptospira interrogans*
9		*Mycoplasma capricolum*
8		*Mycoplasma gallisepticum*
7	√	*Salmonella typhimurium*
7		*Synechococcus* sp
6		*Buchnera aphidicola*
5		*Actinobacillus actinomycetemcomitans*
5	√	*Micrococcus luteus*
5		*Streptococcus pneumoniae*
4		*Chlamydia muridarum*
4		*Mycobacterium bovis*
4		*Pseudomonas putida*
3		*Aquifex pyrophilus*
3		*Campylobacter jejuni*
3		*Vibrio cholerae*

[a] The seven target microorganisms that were cultured for the blind study are indicated.

on the type of the penultimate amino acid. In particular, Met is cleaved if the penultimate amino acid is either Gly, Ala, Pro, Ser, Thr, Val, or Cys. Such proteins have a greater than 50% chance for N-terminal Met loss. This rule was derived from studies on the activity of N-terminal aminopepidases in prokaryotes[14] and accounts, for example, for the bulk of experimentally observed Met losses in the *Escherichia coli* proteome.[8] Since for some SWISSPROT protein sequences the N-terminal Met cleavage has already been incorporated during annotation (PTMs are flagged in SWISSPROT/TrEMBL feature fields), we first restore these Met residues before uniformly applying the deterministic rule to all the biomarker sequences. We would expect that curated PTMs would improve overall system performance. However, in this study the goal is to illustrate the implementation and scalability of an automated modeling and identification system, rather than to subjectively adjust the models to obtain optimal performance.

Identification Algorithm. To identify an unknown microorganism from its mass spectrum, the list of experimentally derived masses is compared to the ribosomal biomarker mass list of each microorganism in the database. A *p*-value is calculated for each microorganism from the observed number of matches and is used to rank each microorganism relative to the others. The identification algorithm selects the microorganism with the smallest *p*-value, provided the smallest *p*-value is less than a Bonferroni-corrected[15] threshold *p*-value of the form $(1 - f)/N$. This threshold *p*-value accounts for the number of microorganisms in the database (N) and the desired confidence level (f). Thus, a database size of 38 and a 95% confidence level corresponds to a threshold *p*-value of 0.0013. The algorithm makes no identification if no microorganism is identified at the 95% confidence level, i.e., if the *p*-values for all microorganisms in the database are above the threshold *p*-value.

Table 2. List of the Microorganisms Cultured for the Blind Study[a]

genus	species	strain	Gram stain	bio-markers	class
Bacillus	*subtilis*	B459	+	31	target
Bacillus	*stearothermophilus*	467	+	20	target
Acinetobacter	*calcoaceticus*	ATCC 19606	−	0	nontarget
Haemophilus	*influenza*	ATCC 9007	−	25	target
Salmonella	*typhimurium*	ATCC 14028	−	7	nontarget
Micrococcus	*luteus*	ATCC 4398	−	5	nontarget
Pseudomonas	*aeruginosa*	ATCC 27853	−	26	target
Escherichia	*coli*	ATCC 25922	−	30	target

[a] *A. calcoaceticus* was cultured as a negative control and is not present in the database (Table 1).

Table 3. Average and Standard Deviation for the Number of Significant Peaks in Mass Spectra for the Eight Species, Obtained with the Four Different Experimental Protocols

ion mode	matrix	trials	⟨peaks⟩	SD
+	CHCA	40	22.9	8.29
+	SA	40	11.5	4.86
−	CHCA	39	11.8	8.39
−	SA	40	10.7	5.24

Experimental Mass Spectra. For the blind study, we cultured eight microorganisms (Table 2), which were intended to represent unknown microorganisms. Five of the cultured microorganisms had models with 20 or more ribosomal protein biomarkers. These were designated as *targets* for identification. Three of the cultured microorganisms had only seven, five, and zero ribosomal protein biomarkers. We did not expect to classify these microorganisms. They were included to test the specificity of our approach and were designated *nontarget* microorganisms.

We introduce variability into the experiment by obtaining mass spectra in two ion polarities and with two different MALDI matrixes, CHCA and SA. A total of 159 spectra were scored (5 replicates for each of the 8 microorganisms in 4 matrix-polarity combinations, with 1 spectrum omitted due to mislabeling). Typical replicate mass spectra, obtained from *Bacillus stearothermophilus* and *Pseudomonas aeruginosa*, are shown in Figures 1 and 2, respectively. These mass spectra typify the qualitative differences between different organisms and protocols. We use the sample mean and standard deviation in the number of significant peaks in a set of mass spectra collected with a particular protocol to quantify the variability of that protocol (Table 3).

(14) Gonzales, T.; Robert-Baudouy, J. *FEMS Microbiol. Rev.* **1996**, *18*, 319–344.

(15) Hochberg, Y.; Tamhane, C. A. *Multiple Comparison Procedures*; Wiley: New York, 1987.

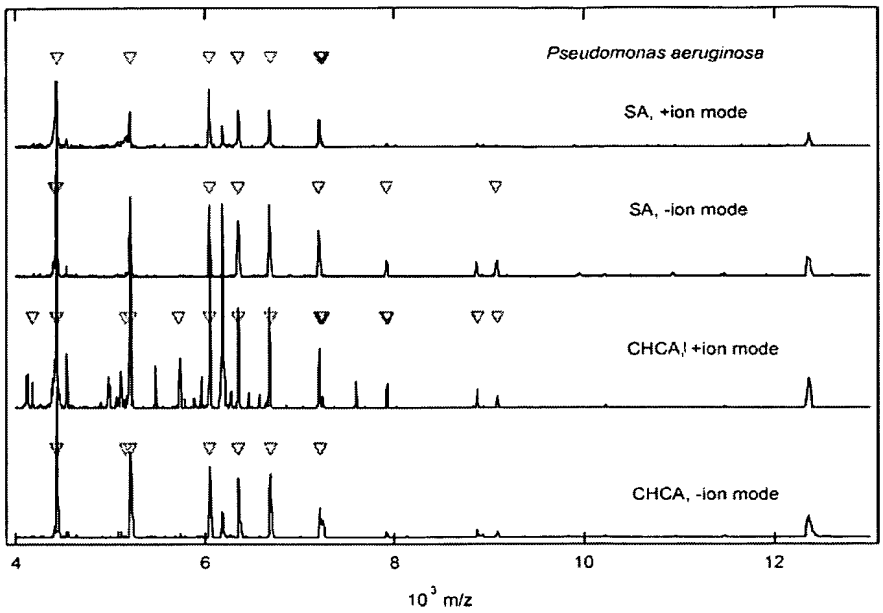

Figure 1. MALDI mass spectra from *P. aeruginosa* obtained with the four different experimental protocols. Peaks that match ribosomal biomarkers (within ±5 Da) are marked.

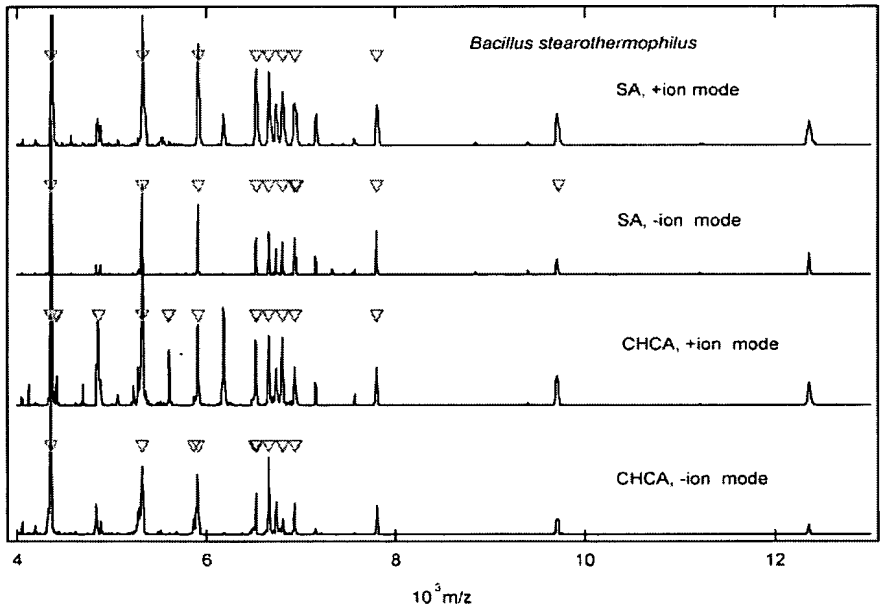

Figure 2. MALDI mass spectra from *B. stearothermophilus* obtained with the four different experimental protocols. Peaks that match ribosomal biomarkers (within ±5 Da) are marked.

Mass spectra obtained with CHCA and in either ion polarity are the most variable—with roughly twice the standard deviation as mass spectra obtained with SA. Positive ion spectra obtained with CHCA exhibit roughly twice as many peaks (compared to the other protocols) due to the well-known tendency of this matrix to generate multiply charged protein ions. Many ions observed in positive mode are not observed in negative mode, suggesting that the former are doubly protonated (see, for example, the 4000–7000 mass range in Figures 1 and 2). The SA matrix is the most reproducible—both the mean and standard deviation of the number of significant peaks are small.

Blind Study Results. The biomarker database and the classification algorithm parameters are fixed independently and prior to scoring of the coded experimental spectra. Over all target/protocol combinations (100 spectra for the 5 target microorganisms) a correct identification rate of 95% with no false identifications is achieved (Table 4). Positive ion spectra yield the best results with a 98% detection rate versus 92% for negative ion mode

Table 4. Identification Rates (%) for Target Microorganisms at the 95% Confidence Level for the Model Biomarker Database with $N = 38$ Microorganisms

species	biomarkers	positive ion mode		negative ion mode		total by species
		CHCA	SA	CHCA	SA	
Bacillus subtilis	31	100	100	100	100	100
Escherichia coli	30	100	100	100	100	100
Pseudomonas aeruginosa	26	100	100	100	60	90
Haemophilus influenzae	25	100	100	80	100	95
Bacillus stearothermophilus	20	80	100	80	100	90
total by protocol		96	100	92	92	95

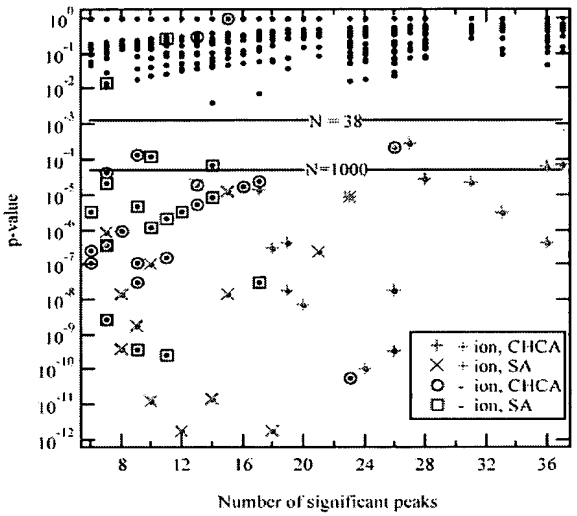

Figure 3. A scatterplot showing *p*-values from each of the 3800 comparisons between microorganism models in the database and experimental mass spectra obtained from target organisms. Points corresponding to correct identifications are further labeled by the experimental protocol. Bonferroni-corrected threshold *p*-values for the 95% confidence level for databases of $N = 38$ and $N = 1000$ microorganisms are marked with lines.

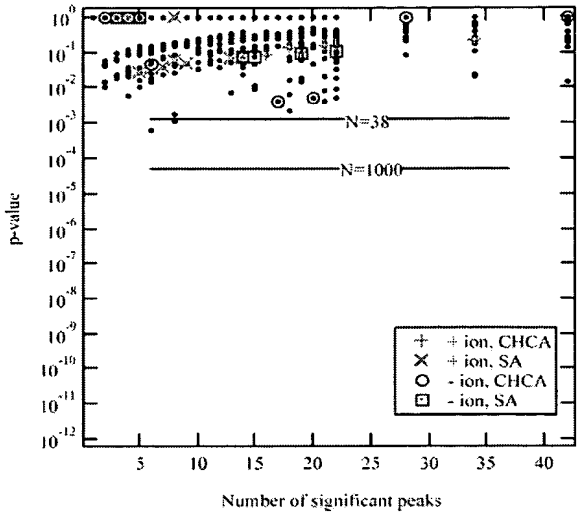

Figure 4. Same as Figure 3, except that the *p*-values come from the 2242 comparisons between microorganism models in the database and experimental mass spectra from nontarget microorganisms. The two points below the $N = 38$, 95% confidence threshold, are false identifications.

(at the 95% confidence level). Detection of 100% is achieved with the most reproducible protocol (SA in positive ion mode). The microorganisms with the greatest number of biomarkers had the best detection rates.

Invariably, six or more significant peaks were found in the experimental mass spectra (with a 2-mV detection threshold). Performance did not depend on the number of significant peaks. A scatter plot of the *p*-value versus the number of significant peaks in the spectrum for each of the 100 experimental target spectra is presented in Figure 3. The bulk of the *p*-values used for correct identifications range between 10^{-4} and 10^{-12}. These are well separated from the *p*-values associated with incorrect identifications. Only 2 out of 59 mass spectra of nontarget microorganisms satisfied the detection threshold (Figure 4). This is consistent with our expectation that nontarget microorganisms would have insufficient biomarkers for robust identification. The two assigned mass spectra were incorrectly assigned to nontarget microorganisms. Had the database included only fully sequenced microorganisms, there would have been no misidentifications among any of the experimental mass spectra.

DISCUSSION

A fundamental advantage of proteome-based methods over fingerprint-based methods for microorganism identification is the mode of database generation and maintenance. Fingerprint methods require the collection of mass spectral fingerprint data to expand their biomarker databases. On the other hand, for proteome database methods, the databases can be expanded automatically as new sequence data become available in the course of genome-sequencing projects. The number of sequenced microorganisms is still modest (~100), but it is increasing exponentially. Moreover, sequencing technologies are evolving rapidly and costs are dropping quickly,[16] so we envision that, within a few years, genome databases will contain most microorganisms that might require rapid identification (e.g., pathogens).

In light of this anticipated wealth of genomic data, it is reasonable to question whether our approach will continue to perform well as databases become more populated. Accordingly, we recalculate the 95% confidence threshold assuming a database with $N = 1000$ microorganisms. The 95% confidence levels for $N = 38$ and $N = 1000$ are marked in Figure 3. Even with this more stringent detection threshold, the detection rate for the well-

(16) Pennisi, E. *Science* **2002**, *298*, 735–736.

characterized microorganisms using the most reproducible protocol (SA in positive ion mode) remains 100% with no additional false detections. Over all protocols, we cannot identify a total of 11 experimental spectra corresponding to an overall detection rate of 89%. The false identification rate remains the same. The decreased overall detection rate indicates a decrease in robustness. Nevertheless, these results strongly suggest that the current approach would still be useful for databases with as many as 1000 microorganisms.

Another relevant question is whether our approach can distinguish between microorganism strains. We previously presented evidence that *H. pylori* 26695 and J99 strains could be distinguished correctly, albeit with relatively low significance, by using the entire-proteome database search method.[11] Reanalysis of the published mass spectra, using only a database with the ribosomal protein biomarkers, yields *p*-values for the two strains, separated by 3 orders of magnitude. Moreover, the *p*-value for the cultured strain (26695) was well below the 95% confidence threshold for a database with 1000 microorganisms. This strongly suggests that we would have correctly distinguished these two strains if we had cultured them for this blind study. Is the ability to distinguish strains a general feature of our approach? A strain contains heritable characteristics that distinguish it from other strains. If these characteristics are not expressed on the ribosomal proteins, then, of course, there is no possibility of strain identification on the basis of ribosomal proteins alone. It is possible that, because of the critical nature of ribosome structure, ribosomal proteins do not vary as much between strains as they do between species. Even if the heritable characteristics for different strains are reflected in the ribosomal proteins, the ability to distinguish the strain depends on the mass shift induced by the relevant polymorphisms.

For the method to scale up to even larger databases, or to reduce the number of errors with the current database size, it will be necessary to reduce the *p*-values of the true positives by either increasing the mass accuracy of the measurements or improving the model used to estimate the *p*-values. Modeling other abundant classes of proteins is a rational next step. With MALDI sample preparation protocols similar to ours, most of the observed peaks in, for example, *E. coli* spectra can be assigned to cytosolic proteins (only about half correspond to ribosomal proteins).[13] Sample preparation protocols that favor surface proteins[2,17] will also require a different class of biomarker models than considered here. We plan to improve further the robustness of this approach by incorporating such models in more powerful and systematic statistical inference frameworks. In addition, the modeling approach can be expanded to accommodate other experimental techniques for microorganism characterization (e.g., chromatography–electrospray MS[18,19]) or for the generation of protein mass databases.[20]

Finally, we have implemented the methods discussed in this paper in Web-based databases. The database used in this study can be accessed at http://infobacter.jhuapl.edu. A more extensive database is available at http://pinedalab.jhsph.edu/microOrgID. Users of either database can submit mass spectra for identification and can view the biomarkers used to make the identification.

ACKNOWLEDGMENT

Applied Physics Laboratory (APL) investigators were supported under U.S. Navy contract N00024-98-D8124. Much of this project was completed while F.J.P. was affiliated with APL.

SUPPORTING INFORMATION AVAILABLE

An Excel spread-sheet containing a list of the mass peaks for the 159 spectra used in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review January 24, 2003. Accepted May 15, 2003.

AC034069B

(17) Madonna, A. J.; Basile, F.; Ferrer, I.; Meetani, M. A.; Rees, J. C.; Voorhees, K. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2220–2229.
(18) Ho, Y.-P.; Hsu, P.-H. *J. Chromatogr., A* **2002**, *976*, 103–111.
(19) Williams, L. T.; Leopold, P.; Musser, S. *Anal. Chem.* **2002**, *74*, 5807–5813.
(20) Wang, Z.; Dunlop, K.; Long, S. R.; Li, L. *Anal. Chem.* **2002**, *74*, 3174–3182.

Supplement

n	cultured	Species
31	√	*Bacillus subtilis*
30	√	*Escherichia coli*
26	√	*Pseudomonas aeruginosa*
25	√	*Haemophilus influenzae*
24		*Borrelia burgdorferi*
24		*Deinococcus radiodurans*
23		*Mycoplasma genitalium*
23		*Mycoplasma pneumoniae*
22		*Chlamydia pneumoniae*
22		*Chlamydia trachomatis*
22		*Helicobacter pylori*
22		*Helicobacter pylori J99*
22		*Mycobacterium tuberculosis*
22		*Rickettsia prowazekii*
21		*Treponema pallidum*
20	√	*Bacillus stearothermophilus*
20		*Synechocystis sp*
20		*Thermotoga maritima*
18		*Aquifex aeolicus*
16		*Streptomyces coelicolor*
15		*Thermus aquaticus*
14		*Mycobacterium leprae*
10		*Bacillus halodurans*
9		*Leptospira interrogans*
9		*Mycoplasma capricolum*
8		*Mycoplasma gallisepticum*
7	√	*Salmonella typhimurium*
7		*Synechococcus sp*
6		*Buchnera aphidicola*
5		*Actinobacillus actinomycetemcomitans*
5	√	*Micrococcus luteus*
5		*Streptococcus pneumoniae*
4		*Chlamydia muridarum*
4		*Mycobacterium bovis*
4		*Pseudomonas putida*
3		*Aquifex pyrophilus*
3		*Campylobacter jejuni*
3		*Vibrio cholerae*

Table 1. The microorganisms included in the database as well as the number of ribosomal proteins in the 4 to 13 kDa mass range. The microorganisms that were cultured for the experiment are indicated with a check mark. In addition we cultured *Acinetobacter calcoaceticus* as a negative control.